First order primal-dual algorithms for total variation methods with applications to image analysis

paper from Antonin Chambolle, and Thomas Pock

Dong Nie, April 25th

Outline

- Background/Motivation
- Algorithm
- Applications
  - ROF
  - TV-L1
- Discussion

Problem Definition

- Primal problem

$$\min_{x \in X} F(Kx) + G(x)$$

- Legendre-Fenchel conjugate

$$F^*(p) = \sup_{x \in X} \langle p, x \rangle - F(x)$$

- Property

$$F = F^{**}(p) = \sup_{x \in X} \langle p, x \rangle - F^*(x)$$

Problem Definition

- Primal problem

$$\min_{x \in X} F(Kx) + G(x)$$

- Dual problem

$$\max_{y \in Y} -\left(G^*(-K^* y) + F^*(y)\right)$$

- Primal-dual problem

$$\min_{x \in X} \max_{y \in Y} \langle Kx, y \rangle + G(x) - F^*(y)$$

Motivation

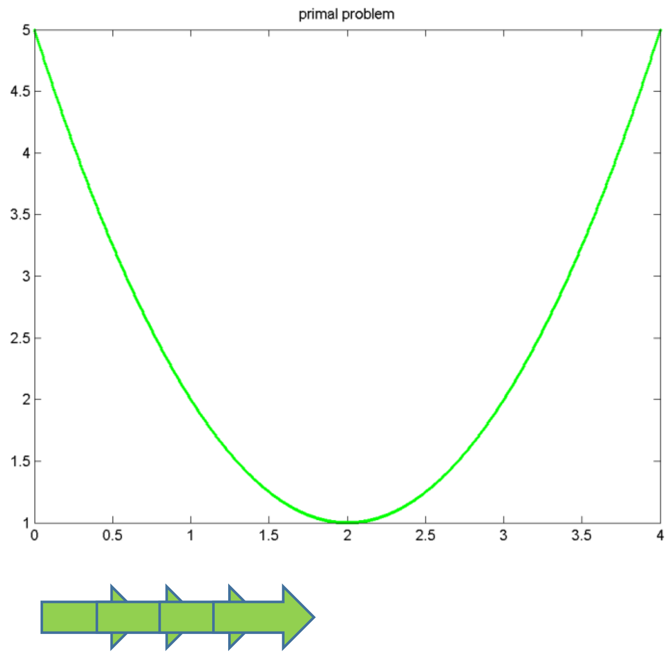
- A lot of convex optimization literature focus on problems where either F or G are smooth, they cannot be applied to where both F and G are non-smooth
- Some methods for non-smooth model is slow
- Sometimes, primal problem or dual problem is difficult to solve
- Most importantly, it supplies more tools to solve convex optimization problems.

Outline

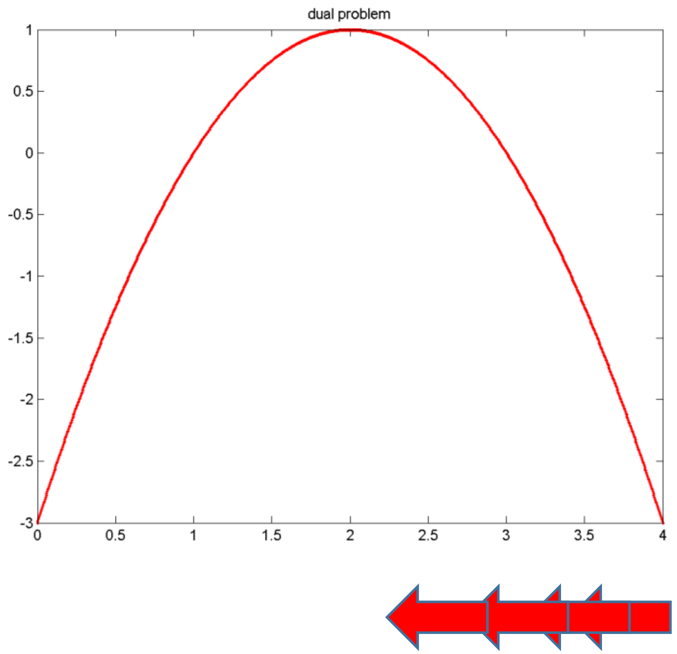
- Background/Motivation
- Algorithm
- Applications
  - ROF
  - TV-L1
- Discussion

Idea

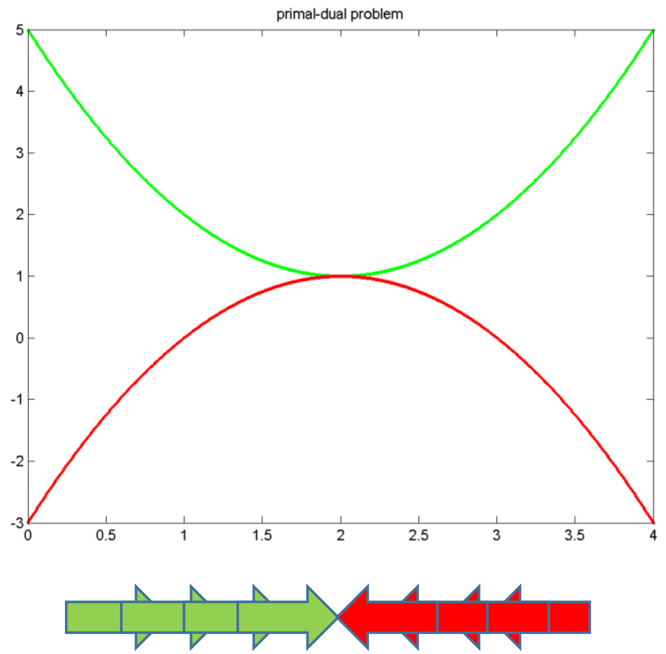
Primal problem



Dual problem



Primal-dual problem



Solution

- $\min_{x \in X} \max_{y \in Y} \langle Kx, y \rangle + G(x) - F^*(y)$
- $\begin{cases} \max_{y \in Y} \langle Kx, y \rangle - F^*(y) - \frac{\lambda_1}{2} \|y - y^n\|_2^2 \to Q(y) \\ \min_{x \in X} \langle Kx, y \rangle + G(x) + \frac{\lambda_2}{2} \|x - x^n\|_2^2 \to P(x) \end{cases}$
- $\begin{cases} (\nabla Q)_y = 0 \Rightarrow y = (I + \partial F^*)^{-1}(y^n + \sigma Kx) \\ (\nabla P)_x = 0 \Rightarrow x = (I + \tau \partial G)^{-1}(x^n - \tau K^* y) \end{cases}$

Algorithm: PDCP1

- Initialization: choose $\tau, \sigma > 0, \theta \in [0,1], (x^0, y^0) \in X \times Y$ and set $\bar{x}^0 = x^0$
- Iterations ($n \geq 0$): Update x^n, y^n, $\bar{x}^n$ as follows:

$$\begin{cases} y^{n+1} = (I + \sigma\partial F^*)^{-1}(y^n + \sigma K\bar{x}^n) \\ x^{n+1} = (I + \tau\partial G)^{-1}(x^n - \tau K^* y^{n+1}) \\ \bar{x}^{n+1} = x^{n+1} + \theta(x^{n+1} - x^n) \end{cases}$$

Acceleration Algorithm: DPCP2

- Initialization: choose $\tau_0, \sigma_0 > 0$ with $\tau_0\sigma_0 L^2 \leq 1, \theta \in [0,1], (x^0, y^0) \in X \times Y$ and set $\bar{x}^0 = x^0$
- Iterations $(n \geq 0)$: Update x^n, y^n, $\bar{x}^n, \theta_n, \tau_n, \sigma_n$ as follows:

$$\begin{cases} y^{n+1} = (I + \sigma_n \partial F^*)^{-1}(y^n + \sigma_n K \bar{x}^n) \\ x^{n+1} = (I + \tau_n \partial G)^{-1}(x^n - \tau_n K^* y^{n+1}) \\ \theta_n = 1/\sqrt{1 + 2\gamma\tau_n}, \tau_{n+1} = \theta_n \tau_n, \sigma_{n+1} = \sigma_n/\theta_n \\ \bar{x}^{n+1} = x^{n+1} + \theta_n(x^{n+1} - x^n) \end{cases}$$

Outline

- Background/Motivation
- Algorithm
- Applications
  - ROF
  - TV-L1
- Discussion

Image Denoising: ROF model

- ROF

$$\min_{x} \int_{\Omega} |Du| + \frac{\gamma}{2} \|u - g\|_2^2$$

- Discrete version

$$h^2 \min_{u \in X} \|\nabla u\|_1 + \frac{\gamma}{2} \|u - g\|_2^2$$

- Primal-dual formation

$$\min_{u \in X} \max_{p \in Y} -< u, \operatorname{div} p >_X + \frac{\gamma}{2} \|u - g\|_2^2 - \delta_p(P)$$

Where $P = \{p \in Y: \|p\|_\infty \leq 1\}$

Image Denoising: ROF model

- ROF

$$\min_{x} \int_{\Omega} |Du| + \frac{\gamma}{2}\|u-g\|_2^2$$

- Discrete version

$$h^2 \min_{u \in X} \|\nabla u\|_1 + \frac{\gamma}{2}\|u-g\|_2^2$$

- Primal-dual formation

$$\min_{u \in X} \max_{p \in Y} -< u, \operatorname{div} p >_X + \frac{\gamma}{2}\|u-g\|_2^2 - \delta_P(p)$$

Where $P = \{p \in Y : \|p\|_\infty \leq 1\}$

In the view of definition:

$$\min_{x \in X} \max_{y \in Y} \langle Kx, y \rangle + G(x) - F^*(y)$$

$F(Kx)\ is \|\nabla u\|_1$ in
ROF and TV-L1 model

$$F^*(p) = \delta_P(p) = \begin{cases} 0, p \in P \\ +\infty, p \notin P \end{cases}$$

Image Denoising: ROF model

- Discrete version

$$h^2 \min_{u \in X} \|\nabla u\|_1 + \frac{\gamma}{2} \|u - g\|_2^2$$

- Primal-dual formation

$$\min_{u \in X} \max_{p \in Y} -< u, \operatorname{div} p >_X + \frac{\gamma}{2} \|u - g\|_2^2 - \delta_p(P)$$

Where $P = \{p \in Y : \forall i, \|p_i\|_\infty \leq 1\}$

$$\|\nabla u\|_1 \Leftrightarrow \max_p < \nabla u, p >, s.t. \|p\|_\infty \leq 1$$

In another view

$$\|\nabla u\|_1 \Leftrightarrow \max_p < \nabla u, p \succ -\delta_p(P)$$

$$\langle \nabla u, p \rangle \Leftrightarrow -\langle u, \operatorname{div} p \rangle$$

Proximity Operator

- Proximity operator of a function

$$(I + \tau \partial F)^{-1}(x) = \underset{y}{\operatorname{argmin}} \frac{1}{2} \|y - x\|^2 + \tau F(y)$$

- For ROF and TV-L1

$$F^*(p) = \delta_P(p) \Leftrightarrow (I + \sigma \partial F^*)^{-1}(p) = proj_P(p) = \frac{p}{\max(1, \|p\|)}$$

$$G_{ROF}(x) = \frac{\gamma}{2} \|u - g\|^2 \Leftrightarrow (I + \tau \partial G_{ROF})^{-1}(x) = \frac{u + \gamma \tau g}{1 + \gamma \tau}$$

Image Denoising: ROF Model

- Apply to ALG1
- $p = (I + \sigma\partial F^*)^{-1}(\tilde{p}) \Leftrightarrow p_{i,j} = \frac{\tilde{p}_{i,j}}{max(1, |\tilde{p}_{i,j}|)}$
- $u = (I + \tau\partial G)^{-1}\boxed{\tilde{u}} \Leftrightarrow \boxed{u} i, j = \boxed{\tilde{u}_{i,j} + \tau\gamma g_{i,j}} 1 + \tau\gamma$
- $$\boxed{\begin{gathered} y^{n+1} = (I + \sigma\partial F^*)^{-1}(y^n + \sigma K\bar{x}^n) \\ x^{n+1} = (I + \tau\partial G)^{-1}(x^n - \tau K^* y^{n+1}) \\ \bar{x}^{n+1} = x^{n+1} + \theta(x^{n+1} - x^n) \end{gathered}}$$

Other Methods

- Gradient Descent (GD)

$$|x| \approx \frac{x^2}{\sqrt{x^2 + \varepsilon^2}}$$

- PDE

$$p_\tau = \nabla u, \|p\| \leq 1$$
$$u_\tau = \mathrm{div} p + \gamma(I - U)$$

- ADMM

$$\min_{u,p,\lambda} \|p\|_1 + \frac{\gamma}{2}\|u - g\|^2 + \langle \lambda, p - \nabla u \rangle + \frac{\alpha}{2}\|p - \nabla u\|^2$$

- …

Different γ

Original image

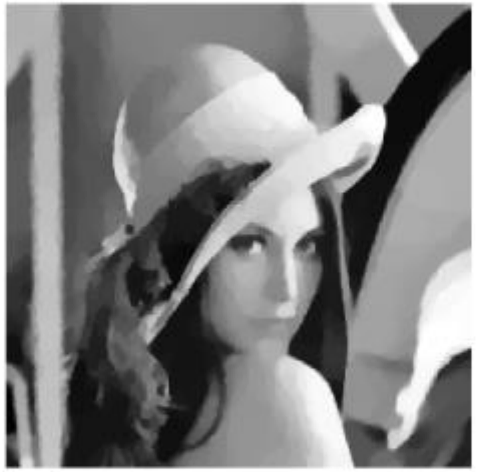
Noise image

$\gamma = 1$

$\gamma = 2$

$\gamma = 4$

$\gamma = 8$

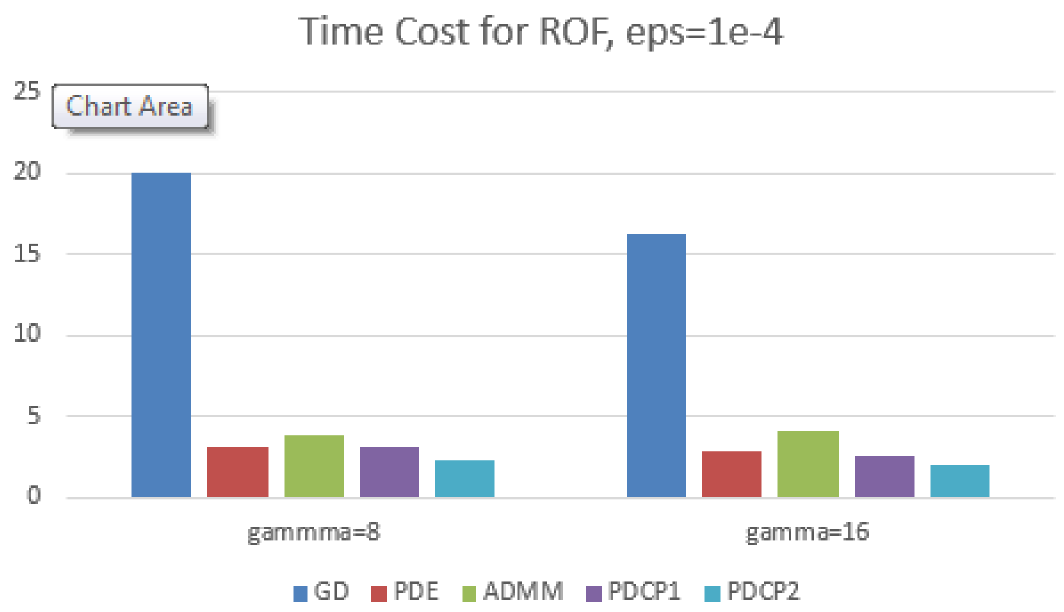


$\gamma = 16$

Different Noise

0.01 0.06 0.1 0.2

Time Cost Comparison



Outline

- Background/Motivation
- Algorithm
- Applications
    - ROF
    - TV-L1
- Discussion

TV-L1 model

- TV-L1

$$\min_{x} \int_{\Omega} |Du| + \gamma \|u - g\|_1$$

- Discrete version

$$h^2 \min_{u \in X} \|\nabla u\|_1 + \gamma \|u - g\|_1$$

- Primal-dual formation

$$\min_{u \in X} \max_{p \in Y} -< u, \operatorname{div} p >_X + \gamma \|u - g\|_1 \; - \delta_P(p)$$

Where $P = \{p \in Y : \|p\|_\infty \leq 1\}$

Proximity Operator

- Proximity operator of a function

$$(I + \tau\partial)^{-1}(x) = \underset{y}{\operatorname{argmin}} \frac{1}{2}\|y - x\|^2 + \tau F(y)$$

- For ROF and TV-L1

$$F^*(p) = \delta_P(p) \Leftrightarrow (I + \sigma\partial F^*)^{-1}(p) = \frac{p}{\max(1, \|p\|)}$$

$$G_{TV-L1}(u) = \gamma\|u - g\|_1 \quad \Leftrightarrow (I + \tau\partial G_{TV-L1})^{-1}(u)$$
$$= \operatorname{shrink}(u, g, \gamma\tau)$$

TV-L1

- Gradient:

$$p = (I + \sigma\partial F^*)^{-1}(\tilde{p}) \Leftrightarrow p_{i,j} = \frac{\tilde{p}_{i,j}}{max\left(1, \left|\tilde{p}_{i,j}\right|\right)}$$

$$u = (I + \tau\partial G)^{-1}(\tilde{u}) \Leftrightarrow u_{i,j} = shrink(u, g, \gamma\tau)$$

$$\text{shrink}(u, g, \tau\gamma) = \begin{cases} u - \tau\gamma, u > g + \gamma\tau \\ g, |u - g| \leq \gamma\tau \\ u + \tau\gamma, u < g - \gamma\tau \end{cases}$$

Different γ

$\gamma = 1$

$\gamma = 2$

$\gamma = 4$

$\gamma = 8$

$\gamma = 16$

Outline

- Background/Motivation
- Algorithm
- Applications
  - ROF
  - TV-L1
- Discussion

Other applications

- TV-based Image problems
  - Image deconvolution and zooming
  - Image inpainting
  - Image segmentation
- More general convex optimization problems
  - Machine learning problems in which loss term and regularization term are both convex
  - Such as SVM, matrix factorization and so on

Discussion

- PDCP is a algorithm framework suitable for problems in which sub-problems are convex, especially non-smooth sub-problems
- PDCP is an example of a first order method, meaning it only requires functional and gradient evaluations.
- PDCP is also an example of a primal-dual method. Each iteration updates both a primal and a dual variable.

Thanks